

# Tarea 1: Probabilidad de bayes y algoritmos geneticos

---

Medel Garduño Diego

## Proyecto 1:

Se implemento un algoritmo genetico sobre datos que representan la cantidad acumulada de personas en un periodo de 1940 a 2023. El algoritmo que se implemento, considera la aplicación de mutaciones efectivas a tres variables importantes en el modelo. Las variables son el número inicial de sujetos al tiempo  $t = 0$ , la tasa de incremento ( $a$ ) y el valor de saturación de la población ( $b$ ), estas mutaciones se consideran efectivas cuando promueven que la curva se ajuste de mejor forma a los datos medidos.

Para resolverlo se llevo a cabo un codigo en python, se considero que las mutaciones hechas al modelo debian ser aplicada al mismo tiempo, es decir, buscar las combinaciones de los tres parametros tales que lograran que la suma de cada una de las distancias individuales entre los datos y los datos del modelo fuera cada vez menor.

## Preguntas:

a) ¿Cuánto valen los parámetros que obtuviste de la funcion logística que mejor ajustan a los datos reales?

El modelo aplicado indica que los valores que mejor ajustan a la curva son los siguientes

$$x_0 = 18.330 \text{ millones}$$

$$a = 0.046$$

$$b = 149.103 \text{ millones}$$

b) ¿Cuántas mutaciones *efectivas* se hicieron en cada parámetro para que el algoritmo converja?

Como se menciona en el preambulo del proyecto, para este caso se busco no aplicar mutaciones individuales, sino que las mutaciones fueran efectivas solo en el caso en que la combinación de valores era tal que la distancia inicial entre los datos y los puntos del modelo fuera disminuyendo.

De tal forma que en el modelo se consideraron un total de 100000 iteraciones y de ese número solo 185 iteraciones fueron mutaciones efectivas

c) Haz la gráfica de los datos reales junto con la curva logística que hayas obtenido que mejor ajusta los datos

Debido al tamaño de la figura, ya que hacerla más pequeña introduce que se pierdan aspectos sobre salientes de la curva, es mejor referenciar en este parrafo. Por ello, para contestar esta pregunta hay que observar la figura 1

d) Si el crecimiento de la población en México continua con la misma tendencia, ¿cuántos años nos vamos a tardar en alcanzar el limite de saturación?

Lo que se observa en la figura 2 es que abra un aumento considerable casi hasta el año 2080, despues de esta epoca se espera que empiece a haber un crecimiento más desacelerado, lo que implica que la curva se acerque a una asintota. Con ayuda del programa realizado se redondeo a tres cifras decimales los valores obtenidos de población, con ello se planteo un ciclo que encontrará el punto exacto donde la curva deja de aumentar y se mantiene igual, para ello se considero una ventana tan amplia de valores, para poder ser cautelosos en si habia un cambio notable en la curva.

El programa arroja que en el año 2223 la curva deja de cambiar, y se mantiene constante en un solo valor, años anteriores se mantiene en los 149 millones de personas, sin embargo, observando los decimales a tres cifras significativas, y analizando el efecto que estas cifras tienen, ya que corresponde a miles de personas, se considero como punto de corte cuando el valor se estipula exactamente en 149.04.

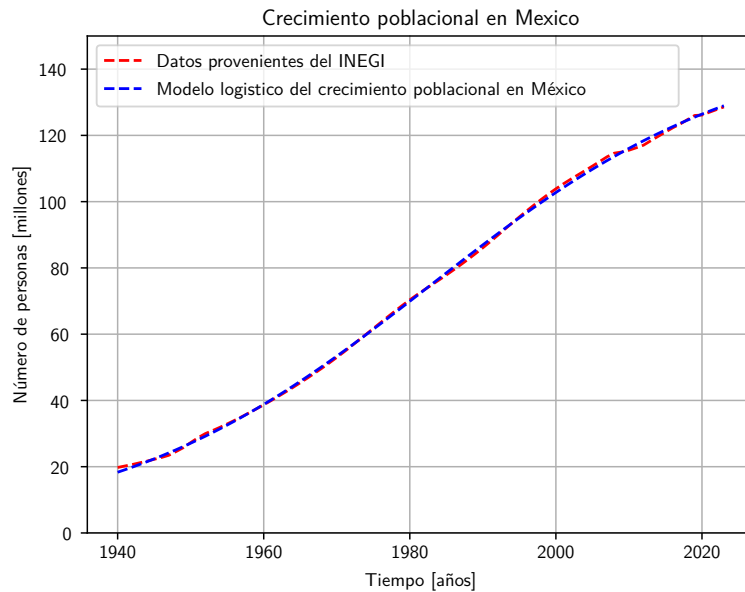


Figura 1: Modelo logístico que mejor ajusta a los datos del INEGI

Entonces vamos a dar dos valores, el primero de ellos indica que la curva deja de cambiar de los 149 millones de habitantes que es en el año 2176, el segundo valor corresponde al punto en el que a dos cifras significativas ya no hay cambio que es 2223.

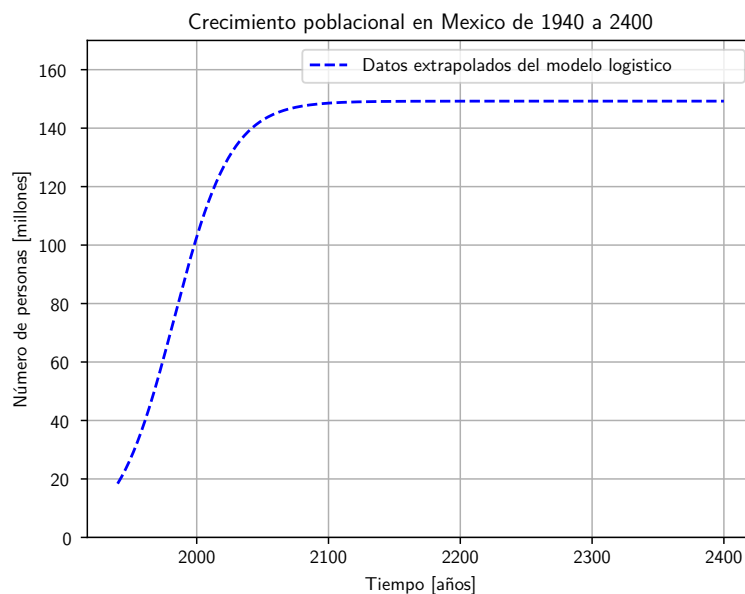


Figura 2: Extrapolación del modelo los próximos 189 años, después del 2024

e) ¿Cuánto vale la población de saturación en México?

Redondeando a un entero, ya que es el valor del que la curva ya no se mueve, tenemos que la población de saturación en México está en los 149 millones de habitantes. Considerando dos cifras significativas el valor de saturación tiene un valor de 149.04 millones de habitantes.

f) ¿Crees que es importante disminuir la tasa de crecimiento poblacional en México?

g) Si consideras que es importante disminuir la tasa de crecimiento poblacional, explica por qué y qué propondrías para hacerlo, Si consideras que no es importante, también explica por qué.

Hay que considerar como primer punto que el máximo de población sensada para el año 2023 es de 128.59 millones de habitantes en México, un país que no ha podido contrarrestar los casos de pobreza extrema y que mucho menos tiene

un plan de contingencia para todos aquellos infantes que tienen que vivir un vida en las calles, donde lo unico que aprenden es hacer uso de sustancias para evitar pensar en el hambre, cortos de oportunidades y posibilidades en un pais como este. Tambien es importante considerar el desabasto de agua, lo caro que es vivir en zonas urbanas como la ciudad de México, la contaminación en las grandes ciudades que cada vez empeoran más el clima.

El modelo planteado indica que estamos a 11.5 millones y 157 años del valor de saturación, lo que deja pensar que si no existe un control en la tasa de crecimiento de la población dentro de esos 150 años que faltan este será un país con sobre falta de oportunidades, no habrá cabida para utilizar automovil propio, el sistema de transporte publico será ineficiente, aun más de lo que lo es hoy si estamos hablando de que 11 millores de personas se sumaran a este país.

Por tanto considero que es importante, sin embargo no tengo una solución que pueda ser viable para disminuir este crecimiento en la población, se supone que al dia de hoy y de forma mundial existe un control de planificación familiar, basado en el uso de anticonceptivos que hasta donde tengo entendido esta al alcance de todos. Sin embargo, hay aspectos que deben ser estudiados y tomados en cuenta, aspectos sociales que esta pasando que cada vez son más los adolescentes que tienen hijos a temprana edad. Por lo pronto, creo que incluso puede llegar a darse el caso de paises orientales que solo se les dada la oportunidad de tener un hijo, alguna institución competente tendrá que plantear un sistema de control natal o tal vez será la unica solución.

## Proyecto 2:

### Preguntas

Para este proyecto es importante establece algunas cosas importantes, ya que a diferencia del otro este proyecto resulto en la obtención de una curva que no converge del todo a la forma de los datos, cosa que puede implicar mucho, en principio lo más cercano es pensar que los datos no se estan comportando de una manera logistica en su totalidad y puede que haya un modelo que ajuste mejor la curva.

En pincipio en la figura 3 se puede observar la manera en la que se comportan los datos acumulados en México de VIH, se escogio que los datos acumulados se encontraran en miles, debido a que es mucho más facil observar su comportamiento.

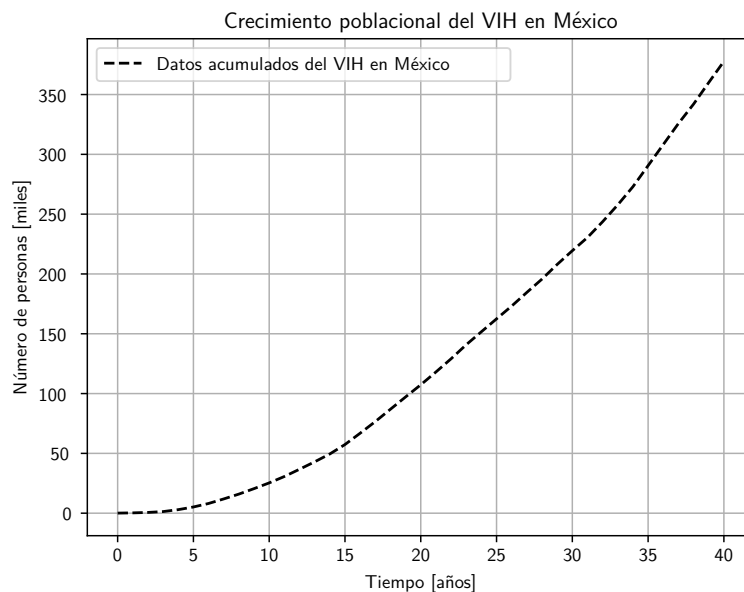


Figura 3: Datos acumulados de VIH en México

Un desafío de este proyecto era encontrar los valores iniciales de las variables  $x_0$ ,  $a$  y  $b$  que mejor se ajustaran a la curva. En la figura 4 se observa el primer acercamiento obtenido, la curva no es muy parecida ya que no tiene la misma tasa de crecimiento que los datos, lo que hace que ascienda de manera más lenta. Por la población inicial se considero el valor de casos para  $t = 0$ , el escoger el valor  $a$  fue más una elección basa en la prueba y error, ya que de manera analitica no se tuvo ningun resultado favorecedor. El valor que de manera visual mejor ajustaba a los datos fue de 0.19, por otro lado  $b$  se escogio considerando el valor maximo de los datos y multiplicandolo por 1.5, lo que dio como valor 453.318 miles de personas.

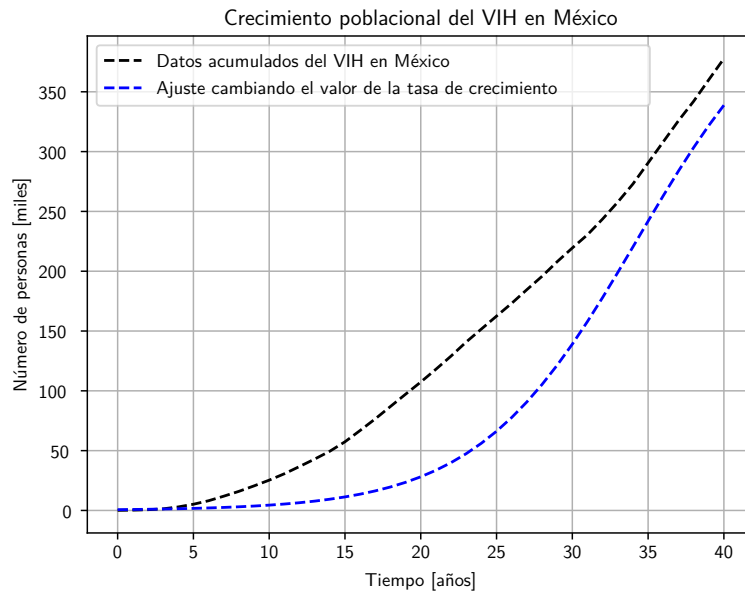


Figura 4: Valores iniciales del modelo logístico

En la figura 5 se observa el modelo que se considero como adecuado, bajo multiples intentos y varios desarrollos este fue la mejor convergencia del modelo, de tal forma que el valor  $x_0$  al que se convergio fue de 8.94, el valor para  $a$  fue igual a 0.130 y por ultimo  $b$  tuvo un valor de 453.318. Con estos valores la curva observada en color azul es muy semejante a la de color rojo, sin embargo hay algunas consideraciones importantes, como por ejemplo que no empiezan en el mismo punto y que la curva azul comienza a descender cuando la curva roja no lo hace.

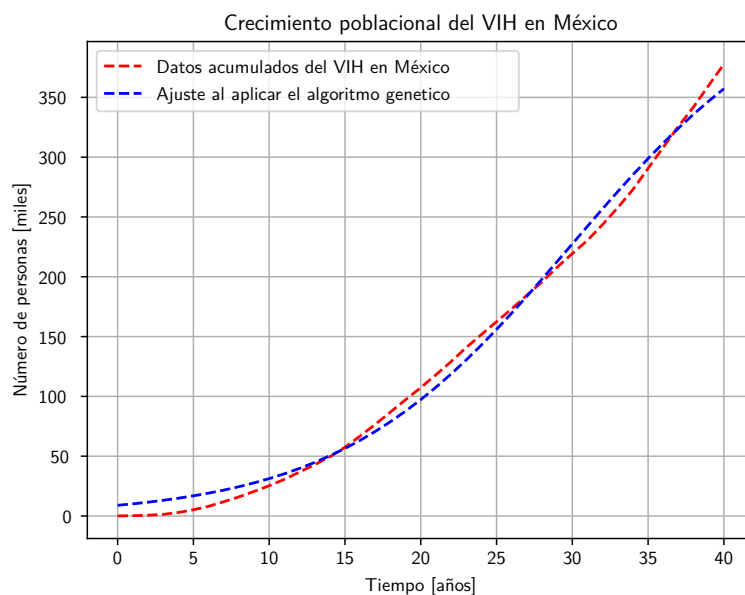


Figura 5: Modelo logístico constratado con la curva de los datos

a) Tomando en cuenta los casos acumulados, si la tendencia sigue como hasta ahora respecto a la población con VIH - SIDA, ¿cuántos años faltan para que la curva se *se aplane*, es decir, para alcanzar la población de saturación?

Se hizo una extrapolación de los datos hasta 2300, se escogio este año con el fin de poder observar de manera completa que es lo que esta pasando con el comportamiento del modelo. Al igual que en la pregunta del homologa a esta pregunta, pero del proyecto 1, se considero que el programa planteado buscará en que momento los datos de la curva comienzan a ser iguales a los datos siguientes. Con ello se observo que para 2100 el valor que se alcanza es de 453 miles de personas con VIH, los subsecuentes años se mantiene en este valor con modificaciones en los decimales. Sin embargo despues del año 2122 los valores del modelo se estacionan en 453.318, es decir, no cambian con tres cifras significativas consideradas.

Contestando a la pregunta y considerando que los datos son hasta 2023, faltan un total de 99 años para alcanzar el punto de saturación

b) ¿Cuánto vale la población de saturación?

La población de saturación tiene un valor de 453.318 miles de personas

c) Con base en las respuestas anteriores, ¿consideras que el VIH - SIDA es una enfermedad que ya paso de moda o que es necesario implementar programas eficientes para atenderla de una mejor manera?

Considerando que el maximo de acumulados hasta 377.765 miles de personas con VIH y el valor esperado de saturación es de 453.318, esto considerando que posiblemente el modelo este siendo minimizando en efecto debido a que no ajusta de la mejor manera a los datos, se puede esperar que la curva se aplane con mas de 100 mil casos de VIH. Esto implica que 100 mil personas serán portadoras de este padecimiento, que limita su calidad de vida y la modifica de manera sustancial. En mi opinion esta no es una enfermedad que ya haya pasado de moda, tal vez es menos critica ya que las personas no se estan muriendo a la misma tasa que en su auge, sin embargo, se esta hablando de que la vida de 100 mil personas se podrá ver comprometida, en ese caso es necesario plantear programas que sean más eficientes en su tratamiento y no solo en el tratamiento, creo que el punto medular no es tratarla, sino evitarla.

Se debe apostar y apuntar al hecho de evitarla de alguna manera, buscar algun tipo de procedimiento que haga que una persona sea inmune a contraer ese virus o en su caso a que no se exprese.

### Proyecto 3:

Sabemos que, en un hospital, el aparato de rayos X para hacer la prueba de este tipo de cáncer arroja un 5 % de **falsos negativos** y un 2 % de **falsos positivos**. Juan hace la prueba y resulta positiva. Juan se espanta muchísimo y íensa que tiene cancer. ¿Será? Definamos los siguientes eventos:

**C = tener cáncer de pulmon ( $\hat{C}$  es no tener cancer de pulmon)**

**X = la prueba salió positiva ( $\hat{X}$  es que la prueba salió negativa)**

Dadas las tasas de falsos positivos y falsos negativos, las probabilidades que conocemos son:

$$P(X|C) = 0.95$$

$$P(X|\hat{C}) = 0.02$$

Lo que Juan quiere conocer es la probabilidad de tener cáncer dado que el examen resultó positivo, es decir  $P(C|X)$ . De acuerdo con la formula de Bayes, esta probabilidad esta dada por:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X|C)P(C) + P(X|\hat{C})P(\hat{C})}$$

Lo que no conocemos a priori  $P(C)$ . Esta es la probabilidad de que Juan tenga cancer sin ninguna información concerniente al aparato de medicion ni de ningun otro tipo

### Preguntas

a) Busca datos en el INEGI, en la Secretaria de Salud publica, o donde quieras, para asignarle un valor a  $P(C)$  suponiendo que Juan es una persona deportista que toda su vida ha cuidado su cuerpo. Calcula entonces la probabilidad  $P(C|X)$ . ¿Debería Juan preocuparse por los resultados del examen?

Para responder esta pregunta hay que comenzar con lo elemental, el desarrollo de esta neoplasia depende de muchos factores de riesgo, es principal es el consumo de tabaco, los subsecuentes tienen una conotación ambiental, como por ejemplo estan expuesto o inhalar radon 222, amilato, ser fumador pasivo, estar expuesto a arsénico, padecer la enfermedad pulmonar obstructiva cronica, la edad, la contaminación ambiental y el más significativo un factor genetico. Dicho lo anterior, los especialistas no discriminan o consideran factor de importancia llevar una vida no sedentaria.

Sin embargo, el gobierno Mexicano se ha encargado de darle al publico algunos resultados interesantes provenientes del Instituto Nacional de Cancer (INCAN). Los cuales indican que en el año 2018 se presentaron 9 mil casos nuevos de neoplasia pulmonar, de los cuales el 85 % se debian al consumo de tabaco. Por otra parte por el proyecto 1, se sabe que la población en México para el año 2018 fue de 124.738 millones de personas.

Podemos hacer uso de la interpretación frecuentista de la probabilidad para saber cual es la probabilidad de que ocurra el tener cancer. Entonces partimos del hecho de que el 15 % de pacientes con cancer de esos 9 mil casos son debidos a no fumar, entonces 1350 pacientes presentaron cancer sin ser fumadores activos, lo que nos lleva a lo siguiente:

$$P(C) = \frac{1350}{124.738 \times 10^6} = 1.082 \times 10^{-5} \quad (1)$$

Sabiendo el valor de tener cancer, podemos obtener el valor de la probabilidad de no tenerlo:

$$P(\hat{C}) = 1 - P(C) = 0.999 \quad (2)$$

Considerando que ya conocemos las otras probabilidades, podemos calcular la probabilidad de tener cancer dado que la prueba salio positiva

$$P(C|X) = \frac{0.95 \cdot 1.082 \times 10^{-5}}{0.95 \cdot 1.082 \times 10^{-5} + 0.02 \cdot 0.999} = 5.142 \times 10^{-4} \quad (3)$$

Dada la condición de no ser fumador, la probabilidad de tener cancer por los otros factores es significativamente pequeña, por lo que posiblemente Juan debería considerar realizar otro estudio o en su debido caso, no estar tan asustado.

**b)** Sebastian se hace la misma prueba y tambien sale positiva. Sin embargo, Sebastian tiene una historia de fumador empedernido desde hace 20 años y de poca actividad física.

Para este caso consideramos el 85 % de los 9 mil casos reportados en el año 2018, entonces en total para personas con cancer que eran fumadoras hay un total de 7650 casos confirmados, por tanto  $P(C)$  es de la siguiente forma:

$$P(C) = \frac{7650}{124.738 \times 10^6} = 6.133 \times 10^{-5} \quad (4)$$

De esta manera, tambien se puede obtener el valor de la probabilidad de ocurrencia del evento complemento

$$P(\hat{C}) = 1 - P(C) = 0.999 \quad (5)$$

Sabiendo esto hagamos el calculo de la probabilidad de tener cancer dado que la prueba salio positiva

$$P(C|X) = \frac{0.95 \cdot 6.133 \times 10^{-5}}{0.95 \cdot 6.133 \times 10^{-5} + 0.02 \cdot 0.999} = 2.908 \times 10^{-3} \quad (6)$$

Este resultado es sumamente informativo, indica que al fumar la probabilidad de contraer cancer de pulmon aumenta en un orden de 5.65 veces frente a no fumar. Aunque la probabilidad de realmente tener cancer dado que la prueba salio positiva, es alarmante observar la diferencia que puede haber tan solo al erradicar un mal habito.

### Referencia para el proyecto 3

La información respecto a los casos de cancer de pulmón en México para el año 2018, fue obtenida de la referencia siguiente:

- Gobierno de México. (s.f.). Tabaquismo y cáncer de pulmón. Comisión Nacional contra las Adicciones (CONADIC). <https://www.gob.mx/salud/conadic/articulos/tabaquismo-y-cancer-de-pulmon>